The Gini Index

Matthew Gherman and Adam Lott (adapted from notes by Robert Brown and Dillon Zhi)

High School I – 10/21/18

In 1905, the economist Max Lorenz introduced an income inequality curve, coined the Lorenz curve. Let the x values between 0 and 1 correspond to the proportion of the population of a given country. On the y-axis, Lorenz placed the proportion of the total income of the population that was received by the bottom x-proportion of the population. In 2006, the highest-earning 20 percent of the American population earned about 60 percent of all income so the bottom 80 percent, represented by x=0.80, received 40 percent. Thus the point $(0.8, 0.4)$ appears on the Lorenz curve. Figure 1 shows a typical Lorenz curve. We also include the line segment connecting $(0, 0)$ and $(1, 1)$. We will call this line the equal distribution line.

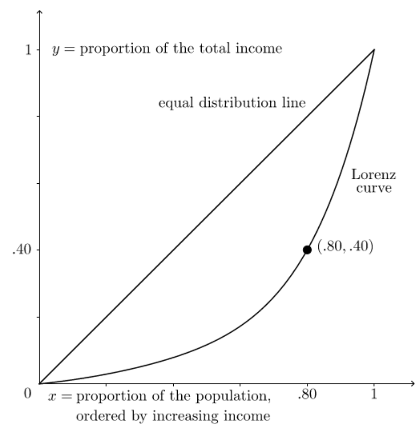

Figure 1:

Exercise 1. Why do the points $(0, 0)$ and $(1, 1)$ lie on the Lorenz curve?

Exercise 2. Explain why the line segment from $(0, 0)$ to $(1, 1)$ should be called the "equal distribution line".

Exercise 3. Why does the Lorenz curve always lie below the equal distribution line? For example, why can't $(.25, .75)$ lie on the Lorenz curve?

The Lorenz curve describes the distribution of income among the members of the community, but how can you compare, for instance, the income distribution of the United States to that of Mexico? In 1912, the Italian statistician Corrado Gini proposed a way to describe the distribution of income by a single number. In Figure 2 the 45-degree line is labeled as (part of) the graph of the identity function $I(x) = x$ and the Lorenz curve is the graph of some function we'll call $L(x)$. The region of the plane between these two curves is labeled by Γ, the Greek capital letter "G", because it was the region of interest to Gini. Since the region beneath the 45-degree line is a triangle with height and base equal to one, and therefore its area is $\frac{1}{2}$, we can see that the area of Γ is no greater than $\frac{1}{2}$.

Exercise 4. What distribution of income would make the area of Γ equal to 1/2?

Exercise 5. Draw an example Lorenz curve where there is large income inequality. Draw an example where there is little income inequality.

In order to present the area of Γ as a proportion of the possible area, that is $\frac{1}{2}$, Gini divided the area by $\frac{1}{2}$, which multiplies it by 2, so it is on a scale that runs between 0 and 1. The result came to be called the "Gini index" so, formally,

$$\text{Gini index} = 2 \cdot area(\Gamma).$$

The region below the Lorenz curve, which we have labeled with Λ, the Greek "L", that is, what is "left over" after taking away the Gini region Γ. We can calculate the Gini index

$$G = 2 \cdot area(\Gamma)$$

if we know the area of Λ because

$$area(\Gamma) + area(\Lambda) = \frac{1}{2}$$

and therefore

$$G = 1 - 2 \cdot area(\Lambda).$$

We can make a rough estimate of the Gini index even from a single observation. Suppose it is estimated that, in some country, the top 20 percent of income earners receive 60 percent of the total income for that country. Thus the other 80 percent of the population shares the remaining 40 percent of the income and we know that the point (.80, .40) lies on the Lorenz curve. Since (0, 0) and (1, 1) also lie on that curve, we'll connect these three points by line segments, as in Figure 3.

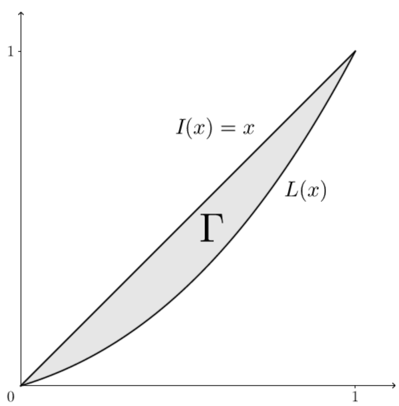

Figure 2:

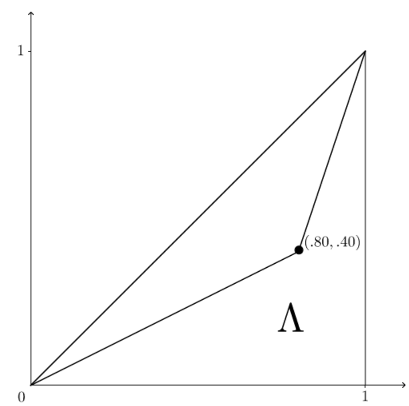

Figure 3:

Exercise 6. Given the Lorenz curve in Figure 3 consisting of two line segments, calculate the area of Λ and use that to compute the Gini index.

In general, if we know that the proportion a of the lowest earners receives a proportion b of the total income, that means that the point (a, b) lies on the Lorenz curve and we can approximate that curve by line segments as in Figure 4.

Exercise 7. Calculate the area of Λ in Figure 4 and then write a general formula for the one-point estimate of the Gini index in terms of a and b.

Exercise 8. Is the one-point estimate larger or smaller than the actual Gini index? Why?

We can get a more accurate estimate of the Gini index if we know two points on the Lorenz curve. In Figure 5 we connected the four points of the Lorenz curve by line segments assuming we knew points (a, b) and (c, d).

Exercise 9. Calculate the area of Λ in Figure 5 and then write a general formula for the two-point estimate of the Gini index in terms of a, b, c, and d.

Exercise 10. Using the points $(a, b) = (.8, .4)$ and $(c, d) = (.99, .8)$, sketch the two-point approximation to the Lorenz curve and use the previous exercise to estimate the Gini index.

Exercise 11. Consider a country with n people. Let the incomes of all the people in the country be $y_1, \ldots, y_n$, sorted in ascending order, and let $Y = y_1 + \ldots + y_n$.

(a) Prove that the Gini index is given by

$$G = \frac{2}{n}\sum_{k=1}^{n}\left(\frac{k}{n} - \frac{y_1 + \ldots + y_k}{Y}\right).$$

(b) Prove that

$$G = \frac{1}{n}\sum_{k=1}^{n}(2k-(n+1))\frac{y_k}{Y}.$$

(c) Prove that

$$G = \frac{1}{2nY}\sum_{i,j=1}^{n}|y_i - y_j|$$

even if $y_1, \ldots, y_n$ are not sorted in increasing order.

(d) The city of Villestown has four people: Alice, Bob, Carl, and Danielle. Their annual incomes are \$20,000, \$40,000, \$60,000, and \$80,000 respectively. Find the Gini index for Villestown.

Exercise 12. Give examples to show that two income distributions which are qualitatively very different can have the same Gini index.

Exercise 13. In Villestown (see Exercise (d)), Alice and Bob form a household and Carl and Danielle form a household. Considering the *household* income distribution rather than the *individual* income distribution, what is the Gini index?

Exercise 14. Give examples to show that even if everyone gets richer, the Gini index may increase, and even if everyone gets poorer, the Gini index may decrease.

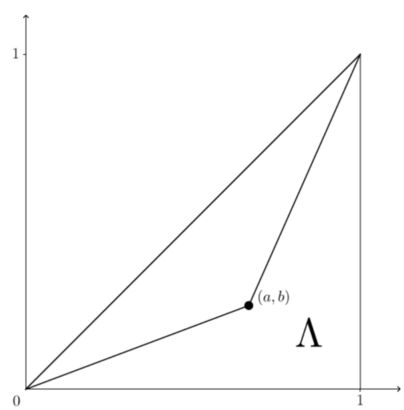

Figure 4:

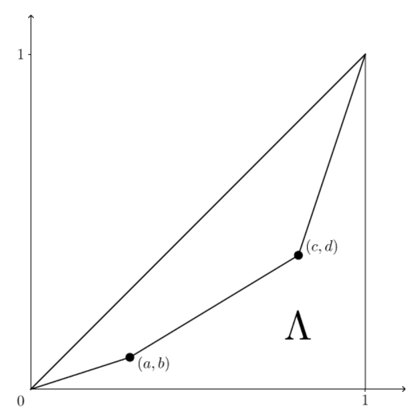

Figure 5: